83Et

United States Department of Agriculture
Agricultural Research Administration
Bureau of Entomology and Plant Quarantine

ELEMENTARY SAMPLING PRINCIPLES IN ENTOMOLOGY

By F. M. Wadley[1]

We usually study insect populations by sampling. A sample may be defined as the portion of a population that is taken for study in the hope that it will be representative enough to tell us what we need to know about the entire population. The material used in an experiment constitutes a sample. For example, the writer reared 53 insects of a certain species to determine the time required for development at 17° C. These insects constituted a sample of the entire population, and from them he hoped to be able to draw conclusions about the species.

Often we wish to study the insect infestation in a plot or field by sampling when the field is itself a sample of the larger population. Then we have compound sampling--that is, samples within samples.

Representativeness

The first thing to keep in mind in sampling is representativeness. We wish our sample to give as accurate an idea as possible of the population under study. We judge representativeness partly by reproducibility. If repeated sampling gives similar results, we believe our samples to be representative.

The best way to make a sample representative is to take it from as many parts of the population as possible. For example, if fruit in an orchard is to be judged from a sample of 1,000 apples, it is better to take 100 apples from each of 10 well-distributed trees than to take 500 apples from each of 2 trees. It is better still to take 10 from each of 100 trees, but the time required to collect the sample may be a practical limitation on subdividing it. Unskilled samplers sometimes think that they can get a representative sample by purposive selection of units that they consider typical. Such sampling is unsafe and may lead to bias.

1/ Formerly statistical consultant of the Bureau of Entomology and Plant Quarantine; now analytical statistician, U. S. Department of Navy.

Freedom from Bias

A second and related principle is freedom from bias. Bias may be defined as a tendency to err persistently in one direction, as does a clock that is always losing time. Objectiveness, or freedom from personal choice, is an important factor in freedom from bias. A man sampling plants in a field to determine the percentage infested or diseased may find that his eye strays subconsciously to plants of the kind in which he is interested. Thus he will tend to get too high or too low a percentage of disease or infestation in his sample. If he notices this, he is likely to adopt some means of taking the choice out of his hands. Several years ago some hessian fly students threw their trowels well into the wheat field and took the strip of drill row nearest the trowel for study. Not all bias is personal. In a set of yield data in individual drill rows of wheat it was noticed that every eighth row was lower in yield than the others. This was believed to be caused by a defect in the drill. A sampling plan based on taking some of every eighth row in this field might have given badly biased results.

Bias is a serious fault in sampling, especially if unrecognized. If our sample is unbiased but not very representative or accurate, we may improve it by making it larger. But if we have a bias, the larger the sample the more definitely will our results point to a false conclusion. A biased sample may be very reproducible. In some sampling work bias is estimated and allowed for, but this is not necessary in most biological problems. Avoidance of bias is our best course, and caution and careful study of each individual problem are the best methods for this avoidance.

Randomness

Randomness is a third principle of importance. It may be defined as giving every unit in the population an equal chance to appear in the sample. Representativeness and freedom from bias are essential for an accurate estimation of the population mean. Randomness insures a good estimate of sampling variation or error. Inferences based on the standard error of the mean apply only to the means of random samples. This is true in general of the inferences we can draw from our statistical tables of reference. Fisher (3) makes clear that randomization is the basis of validity in the error estimate. To calculate an estimate of error it is necessary to have a number of units treated alike; random assignment insures that the estimate is a valid one. Random selection may be achieved by some system of drawing numbers from tables provided for that purpose (see Snedecor 8), or by otherwise carrying out the work so as to insure that each unit has a chance to be represented in the sample. Shewhart (7) gives us valuable methods for use in rather specialized problems, with tests for randomness.

Interrelation of Factors

We are usually interested in central tendencies and variability in a population under study; we may express them as arithmetic mean and standard error, respectively. It is needful to see more of the manner in which factors of representativeness, freedom from bias, and randomness are related to each other and to the average and variation in population.

A random sample is unlikely to be biased. Bias may be avoided, however, without randomness. Randomness and representativeness are somewhat in conflict. Randomness gives us a valid error estimate for the mean. The estimate of the mean itself is of course our primary object. If we desire merely an estimate of conditions at a single time and place, we might even dispense with an error estimate altogether. But if comparisons are to be made with conditions at other times or places, as is usually true, the error estimate also is vital. A few examples will illustrate these points.

Suppose we have a field of drilled corn with 400 rows and about 1,200 stalks per row. We desire an estimate of the average height from a well-distributed sample of 20 plants. We might for some reason believe that the plants on the very edge are not representative and exclude them. This is not a matter of sampling method, but of defining the scope of the inquiry. We could define it as the entire field except for the outer 10 feet, instead of the entire field. We will speak of the entire field here, however. To take a random sample, we might draw numbers for each plant, or obtain them from a table of random numbers. For each plant we would draw two numbers, one from a set of 1 to 400 to indicate the row and the other from a set of 1 to 1,200 to give us the stalk number in the chosen row. Every plant in the field has an equal chance to be chosen at each drawing. It is possible, though not likely, that we would choose the same plant twice. In that case, we would draw another number. This is called sampling without replacement. Such a drawing actually carried out gave the distribution shown in figure 1,a.

The most thorough distribution possible is obtained by spacing the 20 plants in a sort of grid. If there is no basic similarity within rows, 4 rows may be evenly spaced through the field and 5 plants evenly spaced in each row. After we select the starting point, all the plants in the sample are pretty well determined. There is no true randomness; most of the plants have no chance to be included in the sample. The method is free from bias, unless one of the rows chosen falls in a dead furrow or some other such contingency occurs. The distribution is that shown in figure 1,b. This is a purely systematic sample.

With the first method we have unrestricted randomness, and the calculated standard error will give a good idea of the sampling variation of the mean in similar samples and the accuracy of the estimate of the true mean. With the second method we do not have randomness and

cannot correctly calculate sampling error from the sample. If we have several such samples from the same field, with starting points randomized, we can calculate the sampling error among them, but this is not an economical method of work. The even spacing gives the maximum distribution possible in the field. The standard deviation and standard error of the mean will tend to be higher than with a random sample, because these large variations will be fully represented. At the same time the mean will usually be more reproducible on resampling, because by the representative plan each part will appear in every sample. Hence, if we take a single sample by the second method and calculate its standard error as if it were a random sample, we will not get a good measure of reproducibility, but our sample will appear worse than it is.

In some fields there is not much variation between different parts, but a well-mixed variation, so that variation between adjacent stalks is as great as between distant stalks. In such fields random and systematic samples give nearly equivalent results, and error calculated from a systematic sample, as though it were random, is a good indication of reproducibility. But usually differences between parts of the field are more pronounced than differences within parts; and error calculated from a systematic sample, as if it were random, will be too high. If the difference in variation is not very large, there will not be much inaccuracy in this procedure, and what there is will be on the conservative side; but it is not a strictly correct procedure. Efforts are being made to work out methods for calculating an error estimate for a systematic sample, but results so far appear rather difficult for field workers to apply (Madow 6).

Restricted Randomness

A random sample will give an unbiased estimate of the mean which may not be very accurate, and a valid estimate of error. A systematic sample will usually give a better estimate of the mean, but no true error estimate. The systematic sample is the more accurate, because it insures representation of all parts of the field. If we can combine this assurance of representativeness with the valid error estimate, the result will probably be better for our purposes than that by either method discussed. This can be achieved in some measure by restricted randomness. Enough restrictions are laid upon random assignment to insure representation of all important parts of the field.

In our simplified example we might divide the field into quarters and take five stalks at random from each quarter. The mean from the quarters will probably be more accurate than that from a completely random sample, and a valid error estimate can be derived. The standard error of the mean can be computed from the variance within quarters. The variance between quarters need not be used, since in resampling all

quarters would appear each time. In the analysis of variance there would be 3 degrees of freedom between quarters and 16 degrees within quarters. The comparison of the two mean squares will tell whether the modification was helpful. If variance between quarters is higher than that within quarters, the restriction has improved accuracy. The variance of the field mean can be estimated as the within-quarter variance divided by the total number (20), and the standard error as the square root of this value. In the use of the error estimate there will be 16 degrees of freedom. The result of this sort of sampling is seen in figure 1,c. Thus we provide for representativeness and also for our error estimate.

The illustration given is not a very practical one. In practice we would be inclined to take more stalks but in fewer places, because of the time required to get over the field. Sampling evenly through the material is the ideal way, but has practical limits. We might take four or five well-distributed spots with several stalks in each, as a compromise between time saving and statistical efficiency. The variance between stalks within spots would not be a safe guide to calculation of sampling error in such a sample, since variance among adjacent stalks might be low. In this case the spots would really be the sample units, and what we have said about systematic and random samples would apply to them. We might divide the field into quarters, take two spots at random in each quarter and several stalks per spot. In the analysis of variance there would then be three degrees of freedom for variance between quarters and four for variance between spots within quarters. The latter variance would be a sound basis for calculation of standard error of the field mean.

The variance between adjacent units does not provide a valid estimate of sampling error, since it tends to be low, even when the spot is very large. If we should take two entire rows through the field, we should have only two units in our sample in a strict sense.

The example given deals with place-to-place variation in a square field, because this is easy to visualize, but the principles apply to variation other than that of location. In insect-mortality studies we deal with a population varying in resistance from group to group. Resistance may also vary with time of season. The sampling plan should account for all predictable variation by giving it representation, and utilize the uncontrolled variation as an error estimate.

A concrete example may be drawn from insect-population samples taken from one of the uniformity counts of Fleming and Baker (14) by random, systematic, and restricted-random plans, with 50 units in each sample. In this example we have an approach to the actual field conditions under which we must estimate insect populations.

Five samples were taken by each method. A systematic sample was taken by dividing the area of 2,500 units into 50 rectangles, each 5 by 10 units, and taking a unit at the same position in each rectangle. The position was changed for each sample. For the restricted-random

sample plan the area was divided into 25 squares, each 10 by 10 units, and for each sample two units were chosen at random in each square. Purely random samples were taken by methods already discussed.

The five purely random samples of 50 units each gave means and standard deviations as follows:

Sample No.	1	2	3	4	5
Mean	18.5	19.3	20.5	19.3	17.7
Standard deviation	6.3	5.9	7.6	5.5	6.4

The general standard deviation of the total of 250 is 6.4, which is probably a good estimate for the entire population. The mean is a little less than 19.1, whereas Fleming and Baker give the true mean (for all observations) as 19.15. The estimated standard deviation (or standard error) of means of 50 is $6.4/\sqrt{50}$, or 0.9. The actual standard deviation in the five samples we have drawn is computed as 1.0 around the true mean--a good agreement. If we had only one of the samples to use, we should estimate the standard error as 0.9, 0.8, 1.1, 0.8, or 0.9. Thus we can estimate the reproducibility pretty well from a single random sample.

When the five systematic samples were used as though they were random, the means and standard deviations were as follows:

Sample No.	1	2	3	4	5
Mean	19.1	19.4	19.6	18.6	18.4
Standard deviation	6.8	7.3	7.1	7.5	7.0

The pooled standard deviation is 7.2. From the true estimate of random standard deviation, 6.4, we would calculate $s_{\bar{x}}$ as 0.9; from the one calculated from the systematic samples, 7.2, $s_{\bar{x}}$ would be estimated as a little over 1.0. Actually, however, the standard deviation calculated among the five means is much less, a little over 0.5.

This illustrates the tendency of estimated standard deviations of individuals to run rather high in systematic samples, and the means to be more reproducible than with random samples. The systematic samples often provide more accurate estimates than the random ones, but the sampling error cannot be estimated from a single sample of this kind.

Each of the five restricted-random samples was studied by analysis of variance. A typical analysis follows.

	Degrees of freedom	Sum of squares	Mean square
Between blocks	24	1,946.5	81.1
Within blocks	25	869.5	34.8

The standard deviation of random sampling is estimated as $\sqrt{34.8}$, or 5.9; the standard error of a mean of 50 as $\sqrt{34.8/50}$, or a little over 0.8. For the five samples the results are as follows:

Sample No.	1	2	3	4	5
Mean	19.2	20.3	19.4	18.5	18.9
Random standard deviation	4.9	6.7	5.1	5.9	7.0
$s_{\bar{x}}$ estimated	.7	.9	.7	.8	1.0

The pooled standard deviation is a little under 6.0, the estimate of standard error of the means is a little over 0.8, and the actual standard deviation of the five means computed around the true means is about 0.7.

The accuracy of the three methods is reflected in the computed standard deviation of means of successive samples around the true mean--fully random 1.0, systematic 0.5, restricted-random 0.7. It is shown that with the first and third methods a helpful estimate of sampling error can be calculated from a single moderate-sized sample, and that with the restricted-random sample the sacrifice of accuracy is not great.

Often we know in advance something about the population, and can therefore divide the field into rather homogeneous subdivisions, which are superior to arbitrary ones. In the field of figure 1,c, for example, if we divide the area into one subarea of corn known to be tall, one short, and two of medium height, instead of four arranged as square quarters meeting at the center; the subareas may still be equal in size, but might be long and narrow or even irregular in shape.

Devices for Improving Sampling

The restricted-random sampling plan is widely usable. It is often called stratified sampling, and the various subdivisions of the field of inquiry are strata. When the number of units taken in each part is proportional to the size of the parts, the sample is self-weighting. Weighting will be discussed more fully later. A refined mathematical method of determining numbers for each subdivision is to make them jointly proportional to size and standard deviation where the latter is known; that is, proportional to the product of these quantities. Where all standard deviations are similar, the number of units taken in each part is proportional to the size of the part.

In much insect work such choice of size of subdivision sample may lead to oversampling of a part of the field, that is large in size but small in importance, or to the reverse. Perhaps as good a method as any is to sample each subdivision as adequately as possible and combine the results if a general average is needed.

Another sampling device, subsampling or compound sampling, is important and practical in many situations. The major sampling units are not completely studied, but data are determined by subsamples. A familiar illustration is that of estimating the wheat yield of an area by visiting a number of fields and estimating the yield of each one by a moderate-sized sample. We may regard experimental plots as units of a sample, and if the insect infestation of each plot is itself estimated by sampling, we have compound sampling. Analysis of variance applies conveniently to such cases, and by such analysis we can separate the effect of the major and minor orders of sampling on precision. The sampling variance of major units functions as error for questions based on these units; that of minor units within the large groups will be included in the major error.

This sampling device is widely used. In problems of broad scope compound sampling is more usual than simple sampling. In practical work it is not necessary to use randomness in locating minor units within major units, but it is necessary if the minor units are to be used in studying technique. Major units should have some element of randomness, as error estimates are based on them.

Another type of subsampling is the sampling in the laboratory of material gathered as a composite sample from the field. This practice is familiar to chemists and other laboratory technicians. Henderson and McBurnie (15) have described a method of this type of subsampling mite populations on citrus leaves, which reduced labor considerably. In the setting up of such a method, it must be shown that no bias is brought about. Bias may be avoided by using more laborious methods of known exactness as a standard of comparison. Where two or more subsamples are provided for each sample, subsampling variance may be determined.

Henderson and McBurnie also describe mechanical methods of mite collection. Such methods are frequently developed by workers. They are not directly statistical, but are an outgrowth of desire to get the most out of limited time and funds. Sometimes insects can be weighed or measured if an efficient collection procedure is available, and if the relation of the quantity thus determined to actual numbers is established.

The last type of subsampling is a form of double sampling in which the characteristic of interest is hard to measure. We therefore estimate a related characteristic, easier to handle, on a good-size sample, and estimate the relation of the desired characteristic from a more limited sample, with a saving in labor. Double sampling takes various forms. In a study of European corn borer populations, the desired characteristic is borers per 100 plants, but counting is laborious, requiring careful dissection of stalks. Therefore, the percentage of stalks infested is easily estimated on a large sample, and a limited sample is dissected to determine the number of borers per infested stalk. The final figure

is the product of these two. Cases might occur in which the final figure would be a quotient of two variables. In other cases regression of the first character on the second is estimated from a medium-sized sample, the second is then estimated from a large sample, and the final figure is computed from the estimated regression applied to the results of the large sample. This method could be applied when numbers of insects are estimated from measurement or weight.

Double sampling is useful where material has been placed in categories by rapid inspecting methods, as has sometimes been done with number of scale insects on citrus, or amount of damage by earworms to corn. If material in each category is sampled and the samples are used in actual counts, the means of counts may be applied to the categories with improvement in exactness. Estimation can sometimes be considerably improved without a great increase in work. Determination of error in double sampling is complex. Where products or quotients are used, and are calculated separately in every replication or major subdivision, error may be simply calculated among the final figures.

In insect-population work some index of the population is often used rather than an accurate count. Sometimes active or numerous insects are caught with a sweep net, instead of being counted on the plants. Trap catches or screen counts often serve as indices of abundance. Use of such methods assumes a correlation; the correlation must be established if they are to have greatest usefulness. Investigation sometimes shows that sweeping, for example, gives different results on windy and calm days, or that it gives an incorrect picture of sex ratio. These methods are often useful for immediate decisions, but correlation with exact populations must be established if they are to lead to real gains in the knowledge of insect populations.

One method of interpreting sampling results must be condemned. A sampler will sometimes array his data in classes by some qualitative criterion, assign rank numbers to the classes, and proceed to use these numbers as if they were measurements. If infestations are graded as 1, 2, and 3, for example, we have no assurance that 2 is twice as heavy as 1, or that 3 is as heavy as 1 and 2 put together; 2 may be five times as heavy as 1. Double sampling can be applied in this situation with profit.

Some Special Considerations

In sampling in entomology we are generally interested in density of population or in the proportion of the population affected by some characteristic. Both are determined by counting indivisible units, rather than by measuring. This means that sampling variance has a limiting value below which it cannot be expected to go. No amount of precision in procedure will make the variance lower than the minimum value;

for percentage counts, the binomial variance $p.q/n$; for population counts, the mean. The standard error of the mean may of course be reduced by taking larger samples. In low population densities the variance among units is usually close to the theoretical minimum, and in high densities it is greater in proportion to this theoretical value. In low populations sampling error is lower absolutely, and higher proportionally to the mean than in high populations. In percentage counts variance between successive counts is comparatively low near zero and 100 percent, and higher at intermediate values.

A special consideration is that of sampling from a limited population. If the sample makes up a large part of the entire population, it approaches a census. If we measure every plant in a field, we know the average absolutely, without any sampling. If we measure 25 or 50 percent of the plants, the true standard error of the mean will be lower than the classic formula indicates. We can of course estimate the standard deviation accurately from such a large sample. If $\underline{n}$ is the number of units in the sample and $\underline{N}$ the number in the whole field of inquiry $s_{\bar{x}} = (s/\sqrt{n})[\sqrt{1-(n/N)}]$. Using variances as more convenient, we may write $V_{\bar{x}} = (V/n\ [1-(n/N)]$. If $\underline{n}$ is small in proportion to $\underline{N}$, this is the ordinary formula, since $1-(n/N)$ is practically 1. Unless the sample is more than 10 percent of the whole, the adjustment is unimportant. It is of slight importance to entomologists, since our samples are usually small in proportion to our field of inquiry. One entomologist was sampling bark on a large tree for insect infestation and the units were extremely variable. He calculated the standard deviation from a number of units, and attempted to estimate how many units would be required for a desired low standard error, using the equation $s_{\bar{x}} = s/\sqrt{n}$. The answer was absurd, as it indicated that more units must be taken than existed on the tree. The equation $V_{\bar{x}} = (V/n)[1-(n/N)]$ gave a reasonable answer.

Weighting

Weighting in sampling results has caused considerable confusion. The basic principles are as follows: (a) If several parallel samples from the same material are to be combined, weighting by number of units in the sample is appropriate. (b) If the samples represent different parts of a field of inquiry, the best estimate of the average is obtained by weighting by the sizes of these parts. If the parts are equal or nearly so, no weighting is needed. It is assumed that each part is sampled fairly adequately. The mathematical principle of weighting by reciprocal of variance is involved, but need not be developed further here.

As an example of weighting, suppose that three samples, such as those discussed in the section on Restricted Randomness, are taken from the field, each representing all parts of the field. If one is of 50 units and the other two are of 100 each, they should be weighted accordingly. This can be accomplished by adding the totals and dividing by

250 for the mean per unit. If we wish to work with the means per unit, already calculated from the three samples, we can multiply the small sample mean by 50, each of the other two means by 100, add the products, and divide by 250. If we keep the same proportions, we can simplify the multipliers to 5, 10, and 10 and divide by 25, or to 0.2, 0.4, and 0.4 without any division of the sum.

Suppose, however, that the field is of 40 acres, 16 of one soil type and 24 of another. In combining the two samples we give the mean of one part a weight of 16/40, the other a weight of 24/40. The result is our best estimate of the average condition in the entire field.

It is obvious that in the latter case we may be combining things not very similar, and that a more critical procedure would be to state the averages separately. However, we are constantly being called upon for statements of averages, such as the average crop yield for a State, or the average infestation of some insect for a county. Obviously, we will not always know the proper weights, and so must use approximate estimates or assume equality.

Snedecor (8) discusses sampling and standard errors of weighted averages in more detail than can be done here. Where we have several parallel samples from the same material, we are really combining several samples into a single larger sample. The variance and standard error may be calculated as if it were one large sample. Variances, if already calculated, may be pooled.

When several samples from different parts of the material are combined, Snedecor gives us the formula for variance of a weighted average:

$$V_{\bar{x}w} = S[(V \cdot w^2)/K] / [S(w)]^2$$

where V is the variance among individual units in each class, K is the class number, and w is the weight to be used in each. The principle is that, when variances from unlike classes are combined, they should be weighted by the squares of the class weights, instead of being pooled as are those from like classes.

In one case of insect-population sampling, four environments with equal weight had 6 units each, and a fifth had 70 units, but was to be weighted by only 3.5 because of its small area, while the other environments had weights of 6 each. To obtain a weighted average, the statistics are as follows:

Environment	Mean	Number of units (K)	Weight (w)	Variance of units (V)	Vw^2/K	
A	8.3	6	6.0	13	(36 x 13/6)	= 78
B	10.7	6	6.0	7	(36 x 7/6)	= 42
C	7.0	6	6.0	11	(36 x 11/6)	= 66
D	3.3	6	6.0	7	(36 x 7/6)	= 42
E	13.5	70	3.5	35	(12.25 x 35/70)	= 6
Sum	--	--	27.5	--		234

The weighted mean is $[(6.0 \times 8.3) + (6.0 \times 10.7) + \text{----} + (3.5 \times 13.5)]/27.5$, or 8.1. The variance of this mean is the sum of the column Vw^2/K divided by the squared sum of the $\underline{w}$'s

$$234/(27.5)^2 = 0.31$$

Extracting the square root, we obtain $s_{\bar{x}}$ as 0.56 (rounding to 0.6). The mean then is 8.1 ± 0.6. The large variation between environments does not enter the standard error here.

Planning and Interpreting Sampling

In planning a sampling study the first thing to consider is the objective. The information sought should be clearly defined. If we desire merely to record the presence or absence of an insect species, elaborate sampling suited to estimating population density will not be needed. We need only to look carefully in likely places. If we desire to estimate density, however, looking in the likeliest places is almost sure to give too high an estimate. When sampling for density we must inspect both lightly and heavily infested places.

Next we must consider the methods to be used. We should keep in mind the factors of representativeness, freedom from bias, and randomness, with their functions. Efficiently planned sampling will give better figures for the same amount of work and expense, or equally good figures with less work, than poorly planned sampling. We should utilize all available previous information. Our object in quantitative sampling is, first, to estimate the average conditions, and second, to obtain an idea of the variability. If we have some preliminary idea of variability, we can estimate the amount of sampling needed for an estimate of given accuracy. This accuracy can be measured as the standard error of the mean. In the equation $s_{\bar{x}} = s/\sqrt{n}$ we can supply a preliminary estimate of $\underline{s}$, an acceptable figure for $s_{\bar{x}}$, and solve for the $\underline{n}$. The differences measurable or likely to be missed by the sampling can also be defined. If no preliminary estimate of $\underline{s}$ is available, it is often wise to carry on some exploratory work to obtain one. In such work we will be sampling for the standard deviation rather than the mean. With insect-population counts we may always have in mind the minimum standard deviation.

It is often possible to modify the plan of work midway in investigations, if study of early results suggests methods of gaining efficiency. The precision (measured as $s_{\bar{x}}$) of determination of the mean is governed only by the size of the sample ($\underline{n}$) and the variability ($\underline{s}$). The percentage of the entire population in the sample has no great influence. Taking a fixed percentage is not a sound statistical procedure; a 5-percent sample is a better sample in a large population than in a small one.

Whether such devices as stratified sampling, compound sampling, or double sampling will be helpful depends on the nature of the problem. A knowledge of the material to be sampled will aid in efficient stratification. Arbitrary subdivisions can be made if there is no such knowledge, but more efficient work is usually possible if the subdivisions can be made along lines of known variation.

We may have fields within a district as our principal sample units, and small areas within fields as minor or subsample units. The variation of fields within a district is more important than that of units within fields. The degree to which each source of variation contributes to the error of the final results can be evaluated by use of analysis of variance.

A good example is the preharvest estimation of wheat yield in a county, by using 20 fields as units in a sample of the area, and well-distributed but small subsamples in each field. A small sample will give us nearly as good an idea of the yield in a field as a large one. Differences between fields will usually be larger than between units within fields. If we take a very large subsample, or even a complete harvest, of a few fields, we know the situation in those fields very well, but we do not know the county average well, because fields vary. If we take limited subsamples in each of a large number of fields, we get a better estimate of the county average for the same work.

In such a set-up the standard error of the county mean will be estimated by calculating the standard deviation between field means and dividing it by the square root of the number of fields taken in the county. This standard error will include the large field-to-field variation, and will also have a smaller component caused by sampling variation within fields. That is, if sampling variation within fields is absent (if a complete harvest of each was taken), the standard error will be somewhat smaller. By use of analysis of variance we can estimate the error due to within-field sampling, if within-field units as well as fields are taken randomly. The units can be stated in any convenient form, as yield per subsample unit or per acre, in pounds or in bushels. Suppose we have 20 fields and 5 units per field, with results as follows:

	Degrees of freedom	Mean square
Between fields	19	89
Within fields	80	29

From this summary, using the mean square within fields as $\underline{B}$, the mean square between fields as $k \cdot A + B$, where $\underline{k}$ is the number per field (5), we can calculate $\underline{A}$, the variance between fields over and above that within fields, on a unit basis. This is estimated as (89-29)/5, or 12. Variance of the mean for any combination of $\underline{n}$ fields and $\underline{k}$ units per field will be estimated as $A/n + B/nk$. In this case it will be $12/20 + 29/100$, or

0.89, and the standard error will be $\sqrt{0.89}$, or about 0.94. If we have 50 fields with 2 units per field, the expected variance of the county mean will be 12/50 + 29/100, or 0.53, and the standard error about 0.73. For 10 fields and 10 units per field the standard error would be 1.22.

In this manner we can estimate the effect of changes in sampling plan. To spread out sampling will always give a gain if the mean square between fields exceeds significantly that within fields, and if A has a real existence, which is usually true.

The analysis shown can be adapted to the study of small adjacent areas within one field, and thus to comparison of a few large units with a larger number of smaller units. In such a comparison we think of the large units as made up of adjacent smaller units, and make our analysis within and between larger units. If the smaller units completely occupy the larger unit, they are essentially random; the random choice of the larger unit makes them so. By this method 4 or 5 spots in a field, with 25 units per spot, were found to give as precise results as 2 larger spots with 100 units per spot. In orchard sampling for results of spraying, 8 plots of 1 tree each gave as good results as 4 plots of 3 trees each, under conditions of the orchards used.

Labor and other costs must enter into sampling plans. Very fine subdivision of sampling will often greatly increase the labor of covering the ground. Examination of additional units in a spot may add little expense, and it will increase precision somewhat, even though not so much as studying more spots. We must figure, not the lowest standard error possible, but one that will be acceptably low and within our limits of work and expense. Often a compromise can be made and a good plan worked out that will hold down cost and provide for enough exactness. In an elaborate sampling investigation, however, it may pay to subject costs as well as variances to a more exact study. If we have the variance of large sample units and of subsample units, the costs of each type of unit, and the total allowable cost, we can solve for the best combination of n and k in the following equations:

$$V_{\bar{x}} = A/n + B/nk$$
$$T \text{ (total cost)} = n \cdot CD + nk \cdot C$$

Here C represents the direct cost of each subunit, CD the overhead cost of each major unit above subunit costs; C, CD, A, B, and T are fixed; and we solve by calculus for the n and k giving the lowest value for $V_{\bar{x}}$. Tippett (9) treats this problem in his chapter on experiments. In the solution

$$k = \sqrt{\frac{BCD}{AC}} \text{ or } \sqrt{\frac{BD}{A}} \text{ and } n = T/(CD + kC).$$

For a fixed $V_{\bar{x}}$ and lowest total cost, k is the same and n is estimated as $(kA+B)/k \cdot V_{\bar{x}}$. (Davis and Wadley 13.)

Bibliography

Some of the books listed below contain good discussions of sampling principles. Snedecor's text contains many references to sampling, and Chapter 17 is devoted largely to rather advanced sampling principles. Deming (2) gives recent advances. Much recent work is quite complex or is concerned largely with questionnaire methods. Several recent articles are included here for the benefit of persons wishing to pursue the subject further.

General articles

(1) Cochran, W. G.
1939. The use of the analysis of variance in enumeration by sampling. Jour. Amer. Stat. Assoc. 34: 492-510.

(2) Deming, W. E.
1950. Some theory of sampling. 602 pp. New York.

(3) Fisher, R. A.
1949. The design of experiments. 5th ed. 240 pp. London.

(4) Hendricks, Walter
1942. The theory of sampling. U. S. Dept. Agr. in cooperation with N. C. State Col. Agr. and Engin. 122 pp. Raleigh, N. C.

(5) King, A. J., McCarty, D. E., and McPeek, M.
1942. An objective method of sampling wheat fields to estimate production and quality of wheat. U. S. Dept. Agr. Tech. Bul. 814, 87 pp.

(6) Madow, W. G., and L.
1944. On the theory of systematic sampling. Ann. Math. Stat. 15: 1-24.

(7) Shewhart, W. A.
1939. Statistical method from the viewpoint of quality control. 155 pp. U.S. Dept. Agr. Grad. School, Washington, D.C.

(8) Snedecor, G. W.
1946. Statistical methods. 4th ed., 485 pp. Iowa State Col. Press.

(9) Tippett, L. H. C.
1941. The methods of statistics. 3rd ed., 278 pp. London.

Entomological articles

(10) Beall, G.
1939. Methods of estimating the population of insects in a field. Biometrika 30: 422-439.

(11) Bliss, C. I.
1941. Statistical problems in estimating populations of Japanese beetle larvae. Jour. Econ. Ent. 34: 221-232.

(12) Cassil, C. C., Wadley, F. M., and Dean, F. P.
1943. Sampling studies on orchard spray residues in the Pacific Northwest. Jour. Econ. Ent. 36: 227-231.

(13) Davis, E. G., and Wadley, F. M.
1949. Grasshopper egg-pod distribution in the Northern Great Plains and its relation to egg-survey methods. U. S. Dept. Agr. Cir. 816, 16 pp.

(14) Fleming, W. L., and Baker, F. E.
1936. A method for estimating populations of larvae of the Japanese beetle in the field. Jour. Agr. Res. 53: 319-331.

(15) Henderson, C. F., and McBurnie, H. V.
1943. Sampling technique for determining populations of citrus red mite and its predators. U.S. Dept. Agr. Cir. 671, 11 pp.

(16) Jones, E. W.
1937. Practical field methods of sampling soil for wireworms. Jour. Agr. Res. 54: 123-134.

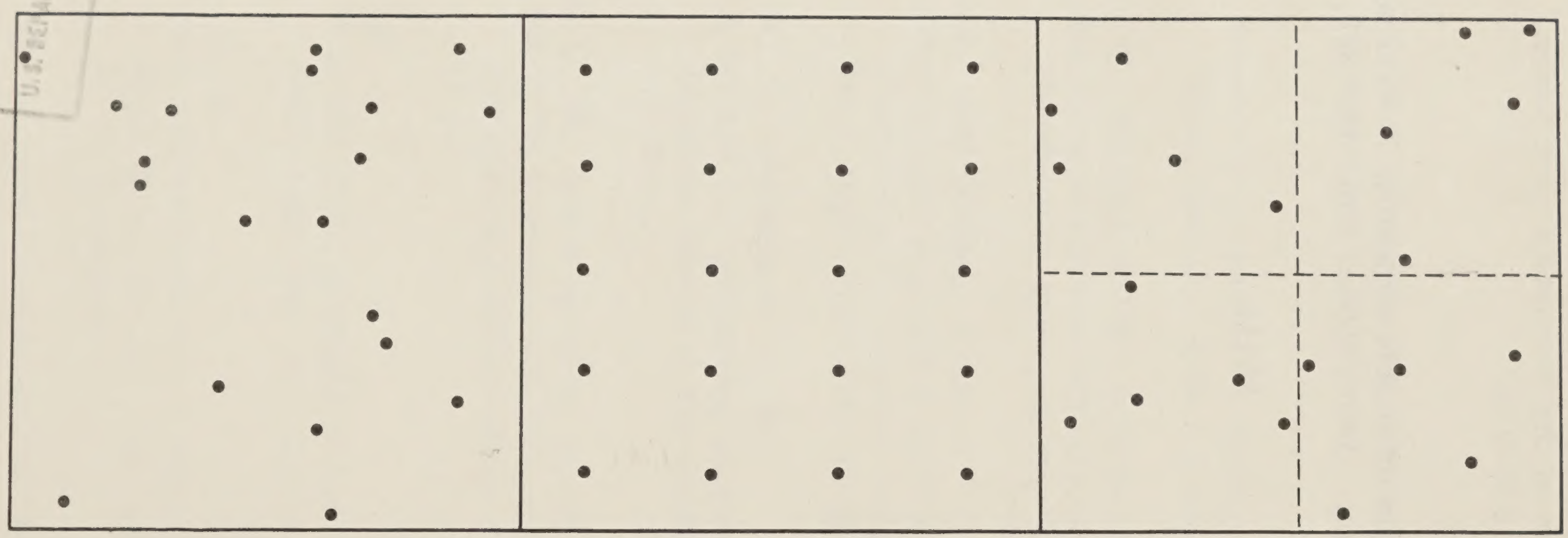
(17) Ladell, W. R. S.
1938. Field experiments in the control of wireworms. Ann. Appl. Biol. 25: 341-389. (Statistical appendix by W. G. Cochran.)

(18) Larrimer, W. H., and Cartwright, W. B.
1926. Determination of the percentage of infestation by the Hessian fly, Phytophaga destructor Say. Jour. Agr. Res. 32: 1049-1051.

(19) Meyers, M. Y., and Patch, L. H.
1937. A statistical study of sampling in field surveys of the fall population of the European corn borer. Jour. Agr. Res. 55: 849-872.

(20) Wadley, F. M.
1949. An application of double sampling in evaluating insect infestation. Jour. Econ. Ent. 42: 396.

§§§§§



a. Fully random sample b. Systematic sample c. Restricted random sample

Figure 1.--Three plans for sampling plants in a field.